ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 10, October 2018

Linguistic Search and Fetching Information with Semantic Annotation

Dr.K.Selvaraj, P.Vidhyavani

Head of the Department, Arignar Anna Government Arts College, Attur, India

Department of Computer Science, Arignar Anna Government Arts College, Attur, India

ABSTRACT: A tremendous growth in the field of mobile computing and increase in deployment of Internet enabled devices in diverse environments has changed the traditional way of computing and communication. With this increasing requirement volatility, software development processes are progressively more relying on reuse of existing softwares. Web services are modular and self-describing applications that can be described, discovered and reused with other Web services to create new software components. But the lack of semantics in the description of services creates inefficiencies in the discovery of potentially relevant components. So, describing the Web service with semantics provides the ability for smart Web service automation. This project proposes semantics based Web services discovery framework for automatic service discovery and execution monitoring using OWL. This framework is mapping the WSDL with OWL-S using Type Convertor and use a Local Ontology Repository to find a suitable ontology for each WSDL specification. Then the Structure, keyword extractions are performed using linguistic search and structural refining methods. The Mobile Web traffic is analyzed using Web Analytics and traffic steering is processed to avoid load overhead. The memory based collaborative filtering algorithm is also used for predictive ranking of services.

KEYWORDS: Web service discovery, Semantic, Annotation, Local Ontology Repository, Linguistic Search, Structural refining

I. INTRODUCTION

Web services are the method of open communication between end to end systems over the World Wide Web. Web service life cycle involves the following phases,

Advertisement: It is the process of publishing the web service in the service registry.

Discovery: It is the process of finding the list of the approapriate web service which provides the user satisfaction.

Selection: It is the process of finding the most suitable Web service in the discovered list.

Composition: It integrates the selected web services into the compound web service.

Invocation: It helps in invocation of web service by providing the needd inputs for its execution.

There are various mechanism involved in the discovery process. Due to the lack of semantics, Web services suffer by means of heterogenities in the form of technologies, ontologies and domain specific tasks. Republishment of web services is often dealt by WSDL due to the dynamic environment in mobile architecture.

To overcome the complexities, Semantic web is used to encode the web service in the machine processable form. It integrates the semantic definition to automate web service discovery process with service extension.

In this approach, the non semantic web services are mapped into the semantic web services using the OWLS-TC type convertor. The convertor is constructed by means of OWLS-TC (Online type convertor) which contains the hierarchial collection of ontologies. These ontologies are describing the input and output concepts of the services in the test collection. The extraction is done in both logical structure and keyword basis. The keyword based extraction is provided by Linguistic search [1]. It extracts the list the keywords used to search in the local ontology repository. These relevant keywords are collected by means of insertion methodology. In insertion, Tokenisation, Lemmatization, Removal of stop words and computing term frequency [10] are the steps involved in the keyword extraction. Then the Ontology Linguistic Relevance (OLR) is calculated to find the optimal result.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 10, October 2018

Structural refining is based on considering the relation among the requested term and local ontology term. After completing the process of checking the ontology list, reorders the ontologies to the computed ranking value. In this approach, "Relation" is the keyword which displays the ontologically crawled web page based on the *synonym, antonym, neighbouring and derived.*

This local ontology store is maintained in the private cloud which personalises the categories of the web page. Then the ranking is based on the search dominant methodology with traffic steering. Website traffic will reduce the throughput of the high quality web pages. Search dominant methodology will enhance the ranking by providing the young high quality web pages. The mobile web search environment is prominent for the information retrieval using ontologies.

This paper is organised as follows Section II and III to demonstrate the background and related work. Section IV denotes the system architecture, Components, Algorithm. SectionV denotes the experimental results. SectionVI concludes the paper.

II. BACKGROUND

2.1. Web Service Description Language

Web services are defined by describing the services as a collection of ports and network endpoints in WSDL. WSDL is the web service description language used to describe the functionality of the service. The WSDL 1.1 is the specification of web service by utilising the http requests (GET and POST) methods. The WSDL 1.2 utilises all the methods in the http request handling. Then WSDL 2.0 is used in RESTfull web services to enhance the stateful transactions by replacing port and port type as endpoint and interface. The functionality of the service is implemented in WSDL by the following categories,

- **Service:** Set of system functions that are registered in the service registry and exposed in the web to provide functionality for the user.
- **Port/Endpoint:** The port is used to establish the connection between service provider and client. It is the aggregation binding needed for the service and its network address. It renamed as <endpoint> in web 2.0 environment.
- **Binding:** It is the interfacing component between the SOAP binding style and transport protocol (SOAP).
- **PortType/Interface:** defines a web service in the form of the operations and the messages. The operation deals with the operations to be performed and the latter deals with the methods that are used to invoke the operations.
- **Operation:** It defines the actions preferred in SOAP and the way of encoding the methods
- **Types:** It describes the data. XML Schema is used inline or referenced for describing the specialization of data.

Although the WSDL provides service description it lacks in semantics. It does not validate the service independently and it is more complex.

2.2. Ontologies

Ontology is a prediction of a shared conceptualization. This prediction is done by means of formal specification os the service, service types and the operations. Ontology is classified into domain and upper ontology. Domain ontology is the conceptualization under the specific category. Upper ontology (knowledge domains) means conceptualise the general concepts of the same domain. In mobile environment, the ontologies are used in the personalisation of the information that is gathered for the users.

OntoLearn system is an infrastructure for automated ontology learning from domain text that uses both natural language processing and machine learning techniques. For given two ontologies, for each concept in single ontology, GLUE helps to find the similarities in the other ontology. Another feature of GLUE uses multiple learning strategies, each of which segregates a different type of information either in the data instances or in the hierarchial structure of the ontologies. It helps in providing the structural refining based on the ontological concepts.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 10, October 2018

2.3. Web ontology language

OWL-S is a representational framework that provides for more complete specifications of the capabilities and behavior of Web Services. It supports greater automation of service-related activities than current representational frameworks for Web Services. It contains three major components. They are service profile, service model and service grounding.

- **Service profile** is used to provide the high level description about the service. It describes the relation of instances in service and profile. It constitutes various parameters to include the service specification.
- **Service model** describes the methodology involved in the service. It comprises the information about what are all the operations involved in the sevice.
- **Service grounding** involves the mapping from the concrete section into abstract section. It mainly deals with the concrete section regarding the transport, serialization and port addressing.

It provides the semantic definition to all the services which helps in automatic web service discovery, web service invocation, Web service composition and interpretation and web service execution monitoring.

III. RELATED WORK

Tamer Ahmed Farrag describes about standardization problems in the dynamic discovery and selection process [1]. In this method, WSDL type to OWL ontology Converter (Type convertor) is used to convert the WSDL typically XSD types to OWL ontologies [5]. The mapping algorithm constitutes the automated phase for extraction of service based on structural refining and Linquistic search [1] and manual phase for defining preconditions and non functional requirements for the OWL-S mapping to achieve QoS parameters. This work does not provide enhanced mobile web search.

T.A. Farrag and H.A. Ali describe a clustering based approach for semi-automatically taxonomy construction for web services that are registered in service repository [4]. The semantics of inputs and outputs of a service are specified in the service profile section of OWL-S description of that service by mapping each input/output parameter to a concept which is defined in OWL ontology. Therefore, in order to calculate the semantic similarity between inputs and outputs of two services the semantic similarity of their corresponding OWL concepts can be measured. They introduce attribute-based clustering and similarity-based clustering as two different approaches for detecting the most similar clusters in the bottom-up clustering method. The main drawbacks are the accuracy needed in clustering and inability to recover from storage corruption.

The ASSAM WSDL annotator is a tool that enables the user to annotate a Web Service using a point-and-click interface. The key feature of ASSAM is the ability to suggest which ontological class to use to annotate each element in the WSDL. The recommendations are based on classification mechanism and machine learning approach [2]. Iterative Classification Ensemble is the basic idea behind the ASSAM based on the intrinsic and extrinsic features. This model contains two separate classifiers to deal with these features. The disadvantages are the atomic operation and unavailability of semantics in the ordered list.

MWSAF (METEOR-S Web Service Annotation Framework) [7] is a semantic web based graphical tool that enables you to annotate existing Web service descriptions with Ontologies described by Jorge cardeso in [3]. It facilitates the parsing of WSDL files into Ontologies. This facilitates the user to semi annotate Web service descriptions using machine learning algorithms [3]. It contains the layered architecture which constitutes the data layer, semantics specification layer, semantics at the registry level. The drawbacks are not measuring or evaluating the tool capability and do not concentrates on the new ontology insertion methodology.

Memory-based collaborative filtering is the most commonly used approach.Lazy learning approach in machine learning terms is used in this filtering method, the preferences (in the form of ratings for content) of a community of users are collected and ranked at a web server. Similarity measure is computed between each pair of users based on the content they jointly rated. Next, recommendations for a particular user can be made by considering users that are similar

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 10, October 2018

to him, and checking for content that they liked but that has not yet been rated by the user or that is not yet in the user's collection. The main drawback is making the prediction slow with consumption of memory and time.

In our proposed approach, the personalisation is done by the following mechanism, one by the use of Private cloud and another by means of hosting search egine.

- Private cloud is used for storing the ontologies from client cache to describe which one is prenominant and to categorise on the basis of personalization [10].
- Hosting Search engine enhances the search by implementing the intelligent classifcation and ranking mechanism [6].

This approach enhances the search mechanism by implementing relation mapping mechanism with the consideration of the structural refining [1].

It provides OWL –Q to implement the quality constraints on the service based on lookup, throughput etc. Traffic steering is the load balancing technique used to increase the throughput.

IV. INTELLIGENT DISCOVERY MECHANISM FOR MOBILE WEB SEARCH

In this mechanism, the mobile web search is embellished by means constructing the mapping algorithm to formulate the semantic web services. It considers the factors like storage corruption [4], bottleneck and scalabilty in ontologies. These are evolved by means of the local ontologies that are formed in the basis of distributed environment and personalisation based on ontologies to provide the enhanced web search.

A. System Organisation

The Mobile web search involves the components such as ontology convertor, personalisation chamber, predefined clusters, Crawler and ontology storage repository which are described deeply in the following section B. The process begins with the input specification. The web service which is needed is requested by the user as the non semantic or semantic web services are mapped into the semantic services.

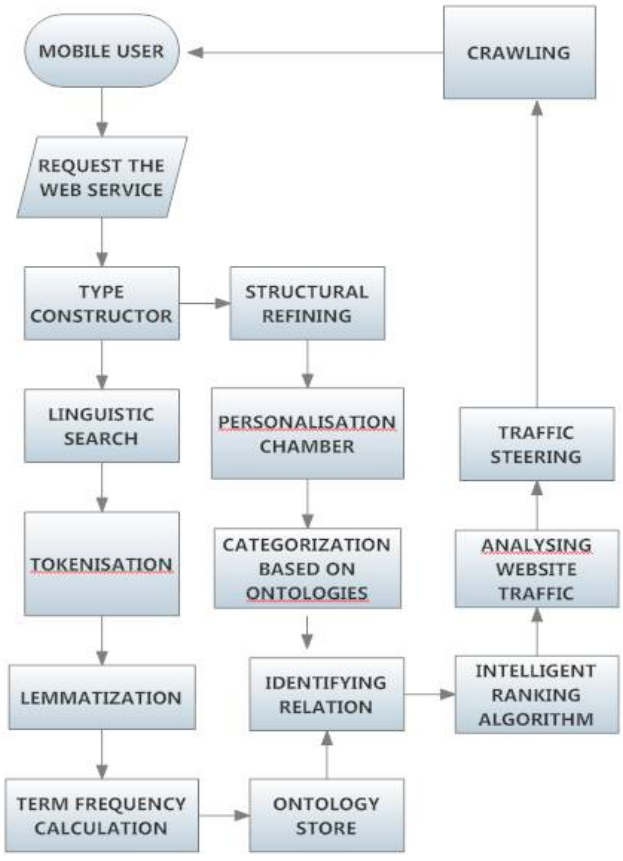

Figure.1. System organisation

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 10, October 2018

Then the extraction mechanism is achieved by keyword extraction based on Ontology Linguistic Retrival to demonstrate the relevant content. The relation is one of the concepts which will display the services based on the relation described for the ontology.

- *Synonym*-exact meaning of the ontology is considered.
- *Antonym*- Opposite meaning of the ontology is considered.
- *Neighbouring*-Relevant concepts are discovered.
- *Derived*- the inherited concepts are discovered.

These are achieved by the structural refining by means of detecting the logical structure.

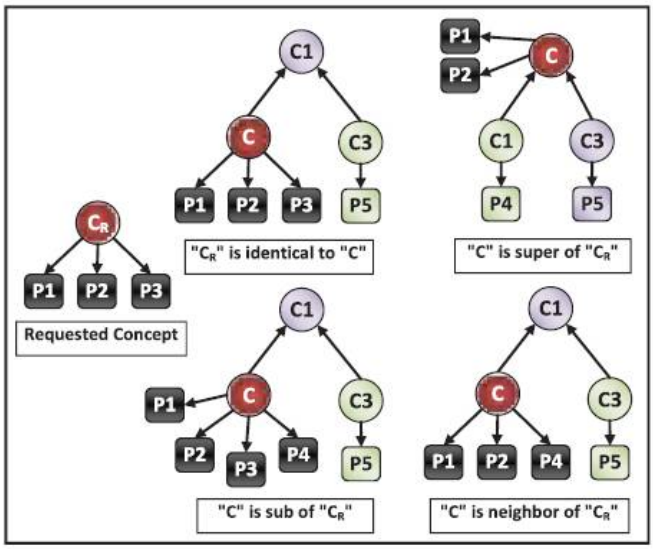

Figure.2. Concept relationships

Due to the insufficient storage in the mobile, we need to implement the private cloud for storage mechanism which will result in the flexible and unlimited computing resource. It provides the private storage area for the specific users. The common data are visible and queried by the other clients except some of the private data. Visualisation is the concept implemented in the storage mechanism because the ontologies are stored in object oriented paradigm.

Authentication mechanism is implemented to access those private data such as business information, mail information etc. Intelligent ranking algorithm is used to provide the effective distance ranking algorithm with search prominent technology. There may be various users accumulated on the single website for searching the information. It provides the load imbalancing in the website which results in the low throughput, high delay and storage corruption. To avoid these problems, Traffic steering method is used to rank the web pages based on website traffic.

This is the basic organisation of the IDM.

B. Components

The discovery mechanism is implemented by the following components.

Type Convertor: The convertor is constructed by means of OWLS-TC (Online type convertor) which contains the hierarchial collection of ontologies. These ontologies are describing the input and output concepts of the services in the test collection. Figure 3 shows a snapshot of the GUI of a tool that is used to insert a new ontology into our repository according to the previously described methodology.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 10, October 2018

Figure.3. Ontology Conversion using OWLS-TC

Ontology store: The database used to store the converted ontologies and prefered for the further references. It acts as a storage chamber in which the converted ontologies are stored in the private cloud for accessing them in mobile environment.

Personalisation chamber: In the personalisation, the ontologies are helpful in providing the categorisation for the users. It helps in providing the categories based on the converted ontologies and their categories.

Predefined clusters: It is used to provide the prefetched services that are registered by converting the previous subsidaries.

Crawler: It is the mechanism at which the web pages are adaptable for the user mobile environment to enhance the user application interface.

These are the components involved in the intelligent discovery mechanism in mobile application.

C. Search dominant distance rank algorithm with traffic steering

This is the efficient page ranking mechanism used in crawling and retrieval of web pages. Distance is defined as the number of clicks provided by the user in that service. The distance between the web pages is calculated as the punishment factor. Although this ranking parameter provides effective on low quality old web pages, the result may be ineffective to the user. So eliminate this problem by means of implementing the search dominant model. In this search dominant model, the pagerank formula is calculated. It is ratio among the numbers of clicks (punishment factor) of the specific time interval.

Algorithm: Crawling algortihm

- The input URL string and the punishment factor is calculated.
- Enqueuing the URL string with the starting_URL.
- If the Queue is not empty, then dequeue the URL and crawl the page of URL.
- If the number of URL queued and Crawled pages are equal, then reorder the queue based on the punishment factor.

Then,

I. Breadth_First {do nothing}
II. Backlink Count {u= number of input links}
III. Page Rank
{

- Calculate the number of links related to the URL and number of links in the page O(P).

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 10, October 2018

- Sort the URL.

}

IV. Distance Rank

Calculate the minimum distance among the user request.

V. Traffic Steering

Calculate the ranking based on the load and website traffic to eliminate the delay time or loading time.

This algorithm is followed by means of search dominant method, it manipulates the punishment factor based on the specific time interval.

Although the search dominant methodology is used the web page traffic must be calculated to reduce the traffic on the single web pages. Then the traffic steering method is used to provide load balancing. Waiting time will negotiate the application functionalities. So the web page with high quality and minimal load will be directed to the web user.

V. EXPERIMENTAL RESULTS

The below figure shows the number of crawled pages are utilised based on the punishment factor. The following graph denotes the bread first search, back link count; Page rank, Distance rank etc. This strategy denotes the throughput of each method.

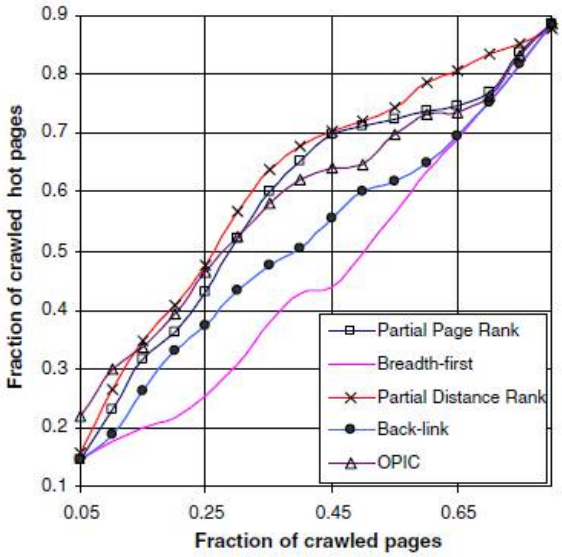

Figure.4. Throughput based on ideal ranking mechanism.

VI. CONCLUSION

The proposed algorithm provides a suitable search on web services over the mobile. This algorithm is a logic matchmaking algorithm with improved results. The structural refining based on the conceptual tree provides the API with enriched capability.

This project show that the algorithm consideration of the standardization problem promises that it will have a positive impact on the discovery process as a whole. The website traffic is analyzed and predicts the ranking by means of predictive intelligent ranking mechanism using search dominant technology.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 10, October 2018

REFERENCES

[1]Tamer Ahmed Farrag, Ahmed Ibrahim Saleh, and Hesham Arafat Ali "*Toward SWSs Discovery: Mapping from WSDL to OWL-S Based on Ontology Search and Standardization Engine* "IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 5, MAY 2013

[2]Heß, A., Johnston, E., Kushmerick, N.: *Assam: A tool for semi-automatically annotating semantic web services.* In: 3rd International Semantic Web Conference, Hiroshima, Japan (2004).

[3] A.A. Patil, S.A. Oundhakar, A.P. Shethand, and K. Verma,"*METEOR-S Web Services annotation Framework*," Proc. 13thInt'L Conf. World Wide Web (WWW), 2004.

[4] T.A. Farrag and H.A. Ali, "*A Cluster-Based Semantic Web Services Discovery and Classification*," Proc. ACME Second Int'l

[5]D. Martin, M. Burstein, D. Mcdermott, S.Mcilraith, M. Paolucci, K.Sycara, D.L. Mcguinness, E. Sirin, and N. Srinivasan, "*BringingSemantics to Web Services with OWL-S*" World Wide Web, vol. 10,no. 3, pp. 243-277, Sept. 2007.

[6] M. Burstein, C. Bussler, M. Zaremba, T. Finin, M.N. Huhns, M.Paolucci, A.P. Sheth, and S. illiams, *"A Semantic Web ServicesArchitecture*," IEEE Internet Computing, vol. 9, no. 5, pp. 72-81,Sept./Oct. 2005.

[7] B. Sapkota, D. Roman, and D. Fensel, "*Distributed Web Service Discovery Architecture*," Proc. Advanced Int'l Conf. Telecomm. andInt'l Conf. Internet and Web Applications and Services (AICT-ICIW '06),2006.

[8]Marc Ehrig and Steffen Staab," *QOM - Quick Ontology Mapping"*, Institute AIFB, University of Karlsruhe.

[9]M. Burstein, C. Bussler, M. Zaremba, T. Finin, M.N. Huhns, M.Paolucci, A.P. Sheth, and S.illiams, "*A Semantic Web Services Architecture*," IEEE Internet Computing, vol. 9, no. 5, pp. 72-81,Sept./Oct. 2005.

[10] Wim F.J. Verhaegh1, Aukje E.M. van Duijnhoven2, Pim Tuyls1, and Jan Korst1, "Privacy protected memory based collaborative filtering algorithm" Springer-Verlag Berlin Heidelberg 2004.

[11] Davide Fossati and Gabriele Ghidoni and Barbara Di Eugenio and Isabel Cruz and Huiyong Xiao and Rajen Subba," *The problem of ontology alignment on the web: a first report"* University of Illinois.